

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
30 August 2001 (30.08.2001)

PCT

(10) International Publication Number
WO 01/62943 A1

- (51) International Patent Classification⁷: C12N 15/64. (74) Agent: HAILE, Lisa; Gray Cary Ware & Freidenrich, 4365 Executive Drive, Suite 1600, San Diego, CA 92121-2189 (US).
C12P 19/34, C12N 15/00
- (21) International Application Number: PCT/US01/05745
- (22) International Filing Date: 23 February 2001 (23.02.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/184,858 25 February 2000 (25.02.2000) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application: US 60/184,858 (CIP) Filed on 25 February 2000 (25.02.2000)
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant (*for all designated States except US*): INVITROGEN CORPORATION [US/US]; 1600 Faraday Avenue, Carlsbad, CA 92008 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): HEYMAN, John [US/US]; 404 Murray Hill Road, Binghamton, NY 13850 (US). SZALAI, Gabor [HU/US]; 1600 Faraday Avenue, Carlsbad, CA 92008 (US).
- Published:
— with international search report
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 01/62943 A1

(54) Title: TOPOISOMERASE LINKER-MEDIATED AMPLIFICATION METHODS

(57) Abstract: A *Vaccinia* topoisomerase-adapted linker, which contains an oligonucleotide primer binding site of known sequence, useful for specifically joining the linker to the end of a polynucleotide of unknown sequence and allowing subsequent PCR amplification of the polynucleotide or DNA is provided. Kits containing the invention *Vaccinia* topoisomerase-adapted linker and one or more linker-specific oligonucleotides for annealing to the linker in PCR are also provided. In addition, the invention provides methods for using the invention linkers in linker-mediated PCR amplification procedures for isolation and optional sequencing of isolated PCR amplification products.

Applicant : Stewart Shuman et al.
U.S. Serial No. : 10/666,486
Filing Date: September 19, 2003
Group Art Unit : Not Yet Known
Title: TOPOISOMERASE-BASED REAGENTS AND METHODS FOR MOLECULAR CLONING

Exhibit 3

26

TOPOISOMERASE LINKER-MEDIATED AMPLIFICATION METHODS

5

FIELD OF THE INVENTION

The present invention generally relates to methods for isolating and sequencing nucleic acid sequences. More particularly, the present invention relates to methods for contiguous sequencing of long DNA containing unknown sequence using a modification of standard PCR techniques.

BACKGROUND OF THE INVENTION

PCR techniques enable the amplification of DNA which lies between two regions of known sequence (K. B. Mullis et al., U.S. Pat. Nos. 4,683,202 and 4,683,195). Oligonucleotides complementary to these known sequences at both ends serve as "primers" in the PCR procedure. Double stranded target DNA is first melted to separate the DNA strands, and then oligonucleotide (oligo) primers complementary to the ends of a target segment whose amplification is desired are annealed to the template DNA. The oligos serve as primers for the synthesis of new complementary DNA strands, using a DNA polymerase enzyme and a process known as primer extension. The orientation of the primers with respect to one another is such that the 5' to 3' extension product from each primer contains, when extended far enough, a segment of sequence that is complementary to the other oligo. Thus, each newly synthesized DNA strand becomes a template for synthesis of another DNA strand beginning with the other oligo as primer. Repeated cycles of melting, annealing of oligo primers, and primer extension lead to a (near) doubling, with each cycle, of

DNA strands containing the sequence of the template beginning with the sequence of one oligo and ending with the sequence of the other oligo.

The key requirement for this exponential increase of template DNA is that the two oligo primers are complementary to the ends of the sequence desired to be amplified, and are oriented such that their 3' extension products proceed toward each other. If the sequence at both ends of the segment to be amplified is not known, complementary oligos cannot be made and standard PCR cannot be performed. Thus, this procedure is impractical for contiguously sequencing a long DNA strand, such as a chromosome. Accordingly, an object of the present invention is to overcome the need for sequence information at both ends of the segment to be amplified, i.e. to provide a method that allows PCR to be performed when sequence is known for only a single region, and to provide a method for the contiguous sequencing of a very long DNA without the need for subcloning of the DNA.

DNA sequencing is a technique by which the four DNA nucleotides (characters) in a linear DNA sequence are ordered by chemical and biochemical means. There are two techniques: 1) the chemical method of Maxam and Gilbert (A. M. Maxam, and W. Gilbert, *P.N.A.S. USA*, 74:560-564, 1977), and the enzymatic method of Sanger and colleagues (F. Sanger, S. Nicklen, and A. R. Coulson, 74:5463-5467, 1977). In the chemical method, the DNA strand is isotropically labeled on one end, broken down into smaller fragments at sequence locations ending with a particular nucleotide (A, T, C, or G) by chemical means, and the fragments ordered based on this information. The four nucleotide-specific reaction products are resolved on a polyacrylamide gel, and the autoradiographic image of the gel is examined to infer the DNA sequence.

In the enzymatic method, an oligonucleotide primer is annealed to a suitable single or denatured double stranded DNA template; the primer is extended with DNA polymerase in four separate reactions, each containing one α -labeled dNTP or dideoxynucleoside-5'-triphosphate (ddNTP) (alternatively a labeled primer can be used), a mixture of unlabeled dNTPs, and one chain-terminating ddNTP; resolving the

four sets of reaction products on a high resolution polyacrylamide-urea gel; and producing an autoradiographic image of the gel that can be examined to infer the DNA sequence. Alternatively, fluorescently labeled primers or nucleotides can be used to identify the reaction products. Known dideoxy sequencing methods utilize a DNA polymerase such as the Klenow fragment of *E. coli* DNA polymerase, reverse transcriptase, a modified T7 DNA polymerase, or the Taq polymerase.

The PCR amplification procedure has been used to sequence DNA being amplified (e.g. using AmpliTaq Cycle Sequencing (Perkin Elmer Cetus Corporation)). By this procedure, DNA can be first amplified and then sequenced using the two conventional DNA sequencing techniques. A modification of this procedure is disclosed by Bevan et al., *PCR Meth. App.* 4:222 (1992)). The PCR method also enables the reduction of non-specific binding of the primers to the template DNA because the enzymes used in these protocols function at high-temperatures, and thus allow "stringent" reaction conditions to be used to improve sequencing.

In the currently existing methods for sequencing DNA of millions of nucleotides, the DNA is fragmented into smaller, overlapping fragments, and sub-cloned to produce numerous clones containing overlapping DNA sequences. These clones are sequenced randomly (sometimes known as the "shot-gun sequencing method") and the sequences are assembled by "overlap sequence-matching" to produce the contiguous sequence. In this shot-gun sequencing method, approximately ten times more sequencing than the length of the DNA being sequenced is required to assemble the contiguous sequence. In these sequencing methods, the linear order of the DNA clones has to be first determined by "physical mapping" of the clones.

Among the currently known contiguous DNA sequencing methods is a procedure called the "primer-walking" or "chromosome walking" method, which uses Sanger's DNA polymerase enzymatic sequencing procedure. In this method, however, the DNA copying always has to occur from the template DNA during DNA sequencing (rather than from the target DNA amplified in the first rounds from the original input template DNA functioning as the template DNA for subsequent cycles

of amplification, as in PCR sequencing). Thus, in the "primer-walking" or "chromosome walking" method, after a certain number of cycles of amplification, the DNA sequencing reaction is initiated by adding a sequencing "cocktail". As a result, the "primer-walking" or "chromosome walking" method requires a larger amount of
5 template DNA than does the PCR sequencing method. Also, when a very long DNA is being sequenced, the DNA has a tendency to re-anneal back to duplex DNA, so that the sequencing gel pattern obtained by the "primer-walking" method may not be as clean as in a PCR procedure. This disadvantage may limit the length of the DNA that can be contiguously sequenced by this method without breaking the DNA.

10 U. S. Patent No. 5,994,058 discloses a method for sequencing long nucleotide molecules using the PCR procedure wherein the sequence of only one primer needs to be known. In this method a first primer is used that is fully complementary to a primer binding site on the target nucleic acid sequence and the second primer consists of 12-16 nucleotides of which 1-10 of the nucleotides anywhere within the primer are
15 of fixed sequence while the remaining nucleotides of the second primer are of random sequence. By generating a large enough number of such second primers of various sequence, one will have a nucleotide sequence fully complementary to a second primer binding site. Using this technique, a long genomic DNA, such as a chromosome, can be contiguously sequenced without the need for subcloning it into
20 smaller fragments. However, in this method, a very large number of second primers must be generated, only a few of which will prove useful.

A combination of "shotgun sequencing" and "chromosome walking" is also currently used to enable the isolation of unknown DNA through use of the adjacent DNA's known sequence. These techniques are routinely applied during analysis of
25 genomic DNA. For example, high-throughput "shotgun sequencing" invariably reaches a stage at which continued sequencing of random clones becomes an inefficient method to close gaps in assembled sequence. When these gaps are too large for standard PCR amplification, "chromosome walking" may be used to systematically obtain and sequence DNA spanning the gaps.

Vaccinia DNA topoisomerase has been used in procedures involving the joining of DNA fragments. *Vaccinia* DNA topoisomerase, a 314 aa virus-encoded eukaryotic type I topoisomerase (I), binds to duplex DNA and cleaves the phosphodiester backbone of one strand (Shuman, S., and Moss, B. (1987) Proc. Natl. Acad. Sci. USA 84: 7478-7482). The enzyme exhibits a high level of sequence specificity, akin to that of a restriction endonuclease. Cleavage occurs at a consensus pentapyrimidine element 5'-(C/T)CCTT-3' in the scissile strand (Cheng, S., et al. (1994) Proc. Natl. Acad. Sci. USA 91: 5695-5699; Clark, J. M. (1988) Nucleic Acids Res. 16: 9677-9686; and Morham, S. G., and Shuman, S. (1992) J. Biol. Chem. 267: 15984-15992). In the cleavage reaction, bond energy is conserved via the formation of a covalent adduct between the 3' phosphate of the incised strand and a tyrosyl residue (Tyr-274) of the protein. *Vaccinia* topoisomerase can religate the covalently held strand across the same bond originally cleaved (as occurs during DNA relaxation) or it can religate to a heterologous acceptor DNA and thereby create a recombinant molecule.

The repertoire of DNA joining reactions catalyzed by *Vaccinia* topoisomerase has been studied in detail by Dr. Stewart Shuman using synthetic duplex DNA substrates containing a single CCCTT cleavage site. When the substrate is configured such that the scissile bond is situated near (e.g., within 10 bp of) the 3' end of a DNA duplex, cleavage is accompanied by spontaneous dissociation of the downstream portion of the cleaved strand (Shuman, S., *J. Biol. Chem.* 267:8620-8627, 1992a; Shuman, S., *J. Biol. Chem.* 267:16755-16758, 1992b). The resulting topoisomerase-DNA complex, containing a 5' single-stranded tail, can religate to an acceptor DNA if the acceptor molecule has a 5' hydroxyl tail complementary to that of the activated donor complex. Sticky end-ligation by *Vaccinia* topoisomerase has also been demonstrated by Shuman, using plasmid DNA acceptors with four base overhangs created by restriction endonuclease digestion.

PCR fragments are naturally good surrogate substrates for the topoisomerase I religation step because they generally have 5' hydroxyl residues from the primers used for the amplification reaction. The 5' hydroxyl is the substrate for the religation

reaction. U.S. Patent No. 5,766,891 discloses a method utilizing this feature of topoisomerase religation to ligate duplex DNAs employing the modified tagged *Vaccinia* topoisomerase. In this method of ligation the donor duplex DNA substrate is a bivalent donor duplex DNA substrate, that is, it contains two topoisomerase cleavage sites. One embodiment comprises cleaving a donor duplex DNA substrate containing sequence-specific topoisomerase cleavage sites by incubating the donor duplex DNA substrate with a sequence-specific topoisomerase to form a topoisomerase-bound donor duplex DNA strand and incubating the topoisomerase-bound donor duplex DNA strand with a 5' hydroxyl-terminated compatible acceptor DNA, resulting in the ligation of the topoisomerase-bound donor duplex DNA strand to the DNA acceptor strand.

Despite these advancements in the art, there is a need for new and better methods for isolating and sequencing long stretches of nucleic acid containing segments of unknown sequence. In particular, there is a need in the art for an efficient method for systematically obtaining and sequencing DNA spanning the gaps in a sequence assembled by high-throughput "shotgun sequencing."

SUMMARY OF THE INVENTION

The present invention overcomes many of the problems in the art by providing an efficient linker-assisted amplification method for isolating and, optionally, contiguous sequencing of an oligonucleotide using basic PCR techniques. The invention is based upon the discovery that *Vaccinia* Topoisomerase I can be used to specifically join a duplex linker containing an oligonucleotide primer binding site to the end of a desired DNA fragment having unknown sequence, allowing subsequent PCR amplification. Invention constructs and methods are particularly useful when the sequence of one primer to be used in the PCR amplification is known and nothing is known about the sequence needed for the other primer. Thus, the invention provides methods and constructs useful for systematically obtaining and sequencing DNA, such as that spanning the gaps in a sequence assembled by high-throughput "shotgun sequencing."

In one embodiment according to the present invention, there are provided double stranded topoisomerase-adapted linkers for use in linker-mediated PCR amplification methods. The invention topoisomerase-adapted linkers comprise a duplex oligonucleotide linker wherein a first oligonucleotide is annealed to a second
5 oligonucleotide that is phosphorylated at the 5' end thereof and wherein the linker further comprises a compatible site-specific topoisomerase enzyme covalently attached to the first oligonucleotide at a single base 3' T overhang.

In another embodiment according to the present invention, there are provided methods for isolating a target polynucleotide having a segment of unknown sequence
10 when sequence of adjacent nucleic acids is known. The invention method comprises cutting the target polynucleotide with a restriction endonuclease that leaves a 3' overhang of at least 2 bases to obtain an overhang-digested polynucleotide segment, dephosphorylating the 5' ends of the overhang-digested polynucleotide segment, performing a Taq polymerase mediated primer extension of the 5' dephosphorylated
15 overhang-digested polynucleotide segment using a sequence-specific primer to create a duplex extension product having at the 3' ends a single A base overhang, incubating the duplex extension product with an invention topoisomerase-adapted linker so as to selectively attach the duplex DNA of the linker to the extension product to form a linked extension product, and isolating a second extension product produced by
20 amplifying the linked extension product using the sequence-specific primer and a topoisomerase linker-specific primer. Since a sequence-specific primer is used in PCR amplification of the overhang-digested polynucleotide segment, it is generally preferred to cut the target polynucleotide in an segment of the oligonucleotide of known sequence.

25 The sequence-specific primers used in the invention methods should be oligonucleotides designed to anneal to the target polynucleotide as close to the unknown sequence as possible, they should not overlap, and should have a melting temperature of at least 56°C so that stringent hybridization conditions can be used during PCR reactions.

In another embodiment according to the present invention, there are provided methods for contiguous sequencing of genomic DNA having a segment of unknown sequence in a vector. In this embodiment, the invention method comprises cutting the nucleotide sequence of the vector with a restriction endonuclease that leaves a 3' overhang of at least 2 bases to obtain an overhang-digested polynucleotide segment that contains an oligonucleotide of unknown sequence, dephosphorylating the 5' ends of the overhang-digested polynucleotide segment, performing a Taq polymerase mediated primer extension of the 5'-dephosphorylated overhang-digested polynucleotide segment using a vector sequence-specific primer to create a duplex extension product having at the 3' ends a single A base overhang, incubating the duplex extension product with an invention topoisomerase-adapted linker so as to selectively join the duplex DNA contained in the linker to the extension product to form a linked extension product, and isolating a second extension product produced by amplifying the linked extension product using the vector sequence-specific primer and a topoisomerase linker-specific primer.

In another embodiment according to the present invention, there are provided kits for contiguous sequencing of genomic DNA that is cloned in a vector. The invention kit comprises an invention topoisomerase-adapted linker and a topoisomerase linker-specific oligonucleotide, wherein the linker-specific oligonucleotide hybridizes with the duplex DNA of the linker under stringent conditions.

In another embodiment according to the present invention, there are provided methods for amplifying by the polymerase chain reaction (PCR) a portion of a target polynucleotide, said method comprising cutting the target polynucleotide with a restriction endonuclease that leaves a 3' overhang of at least 2 bases to obtain an overhang-digested polynucleotide segment, removing the 5' phosphate groups from the overhang-digested polynucleotide segment, performing a Taq polymerase-mediated primer extension of the overhang-digested polynucleotide segment using a sequence-specific primer to create a duplex extension product having at the 3' ends a single A base overhang, incubating the duplex extension product with a

topoisomerase-adapted linker as described herein so as to covalently attach the duplex DNA of the linker to the extension product to form a linked extension product, and performing PCR amplification of the linked extension product using the sequence-specific primer and a topoisomerase linker-specific primer.

5

BRIEF DESCRIPTION OF THE FIGURE

Figures 1A and 1B are schematic illustrations of invention linker duplex oligonucleotides before (Figure 1A) and after (Figure 1B) covalent attachment of *Vaccinia* Topoisomerase I. Names of oligos used to create the pro-linker are in bold. Topo5 (sequence in bold italics) (SEQ ID NO: 1) anneals to Linktop4bio (SEQ ID NO: 2) as shown. The arrow in Figure 1A indicates the break point in the pro-linker at which there is no bond between Topo5 and Linkbot4 (SEQ ID NO: 3). After topoisomerase I makes its covalent attachment to the 3'T residue in the topoisomerase cleavage site (in bold), the double-stranded DNA 3' of this site is the leaving group. LinkAmp4 (sequence shown in italics) and Link Amp5 (underlined) anneal to Linkbot4 as shown. Topoisomerase I covalently attaches to a single base 3' overhang at one end of the cleavage site to create the invention linker topoisomerase-adapted linker as shown in Figure 1B.

10

15

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the present invention, there are provided double stranded topoisomerase-adapted linkers for use in linker-mediated PCR amplification methods. The invention topoisomerase-adapted linkers comprise a duplex oligonucleotide linker wherein a first oligonucleotide is annealed to a second oligonucleotide that is phosphorylated at the 5' end thereof and wherein the linker further comprises a compatible site-specific topoisomerase enzyme covalently attached to the first oligonucleotide at a single base 3' T overhang.

20

25

The site-specific topoisomerase enzyme contained within the invention linkers can be any site-specific type I topoisomerase. Topoisomerases are a class of enzymes that modify the topological state of DNA via the breakage and rejoining of DNA

strands. For example, *Vaccinia* topoisomerase enzyme is a 314 aa type I topoisomerase encoded by a *Vaccinia* virus. Site-specific type I topoisomerases include, but are not limited to, viral topoisomerases such as pox virus topoisomerases. Examples of pox virus topoisomerases include Shope fibroma virus and ORF virus.

5 Other site specific topoisomerases and the cleavage sites to which they bind are known to those skilled in the art.

A "compatible site-specific topoisomerase enzyme," as the term is used herein with reference to the invention linker, means a topoisomerase enzyme that recognizes the cleavage site in the invention linker. Preferably, the invention linker comprises

10 the cleavage site for *Vaccinia* topoisomerase and the site-specific topoisomerase enzyme is *Vaccinia* topoisomerase.

Vaccinia topoisomerase binds to duplex DNA and cleaves the phosphodiester backbone of one strand while exhibiting a high level of sequence specificity, cleaving at a consensus pentapyrimidine element 5'-(C/T)CCTT (SEQ ID NO: 4), or related

15 sequences, in the scissile strand. Examples of alternative *Vaccinia* topoisomerase cleavable sequences are disclosed in U. S. Patent No. 5,766,891, which is incorporated herein by reference in its entirety, and include duplex

5'-GCCCTTATTCCC (SEQ ID NO: 5), duplex 5'-TCGCCCTTATTC (SEQ ID NO: 6), duplex TGTCGCCCTTAT (SEQ ID NO: 7), and duplex GTGTCGCCCTTA

20 (SEQ ID NO: 8).

The phospho-tyrosyl bond between the DNA and enzyme can subsequently be attacked by the 5' hydroxyl of the original cleaved strand, thus reversing the reaction and releasing the topoisomerase. Because topoisomerase catalyzes both cleavage and religation, the structures in Figures 1A and 1B will reach an equilibrium. To assure

25 covalent attachment of the *Vaccinia* topoisomerase to the duplex DNA during formation of the linker (and prevent religation of the cleaved strand), the 5' end of the second nucleotide in the duplex strand of the pro-linker (shown in Figure 1A) is phosphorylated at the 5' end thereof, driving the reaction towards the cleaved product. Once the *Vaccinia* topoisomerase enzyme is covalently attached to the linker and the

leaving group is separated from the pro-linker, the reaction is virtually quantitative and irreversible until an acceptor DNA is provided (i.e., a duplex DNA having a single 3'A base overhang).

A "pro-linker" as the term is used herein, refers to an annealed duplex
5 oligonucleotide (e.g., DNA) substrate that a compatible site-specific Topoisomerase I can cleave and to which the Topoisomerase I will covalently attach at the point of cleavage to yield the invention topoisomerase-adapted linker. Therefore, the pro-linker is a duplex oligonucleotide wherein the first oligonucleotide contains the
10 topoisomerase cleavage site at or near the 3' end thereof, depending upon the preference of the particular topoisomerase used, and the second oligonucleotide is adapted to force the reversible reaction of the topoisomerase with the pro-linker so as to yield the topoisomerase-adapted linker. For example, if the topoisomerase to be used in fabrication of the invention topoisomerase-adapted linker is *Vaccinia*
15 topoisomerase, the cleavage site (SEQ ID NO: 4) is located within about 2 to about 12 bp from the 3' end of the first oligonucleotide in the pro-linker and the second oligonucleotide in the pro-linker is phosphorylated at the 5' end thereof to prevent religation of the leaving group to reform the pro-linker duplex DNA. The method of making the invention topoisomerase-adapted linker is fully illustrated in Example 1
20 herein. Preferably, the first oligonucleotide in the invention pro-linker will contain a sequence comprising the cleavage site for a Topoisomerase I within about 2 to 12 bases of the 3' end.

Incubation of the pro-linker with the compatible site-specific topoisomerase under suitable conditions, as known in the art, will cause the enzyme to cleave the
25 duplex DNA of the pro-linker at the cleavage site, and covalently attach to the 3' end of the cleavage site therein, forming the invention topoisomerase-adapted linker. Other than these requirements, there is no restriction placed on the number or composition (i.e. nucleotide sequence) of the two oligonucleotides in the linker, except that they must be selected so that the two oligonucleotides will anneal and remain annealed during attachment of the topoisomerase to the pro-linker. Preferably

the two oligonucleotides in the duplex linker are fully complementary and the melting temperature of the linker is at least about 56°C.

Thus, the length and nucleotide composition of the two oligonucleotides in the invention linker can be selected for convenience in avoiding unwanted effects that might result from incorporating into the linker an undesirable endonuclease site. However, it should also be kept in mind, that, when transferred to the acceptor end of DNA by action of the enzyme, the linker provides a target annealing site for the "second primer" (i.e., the linker-specific primer) for purposes of performing PCR amplification of a the linked segment of DNA. Consequently, all or any part of the oligonucleotides in the invention linker can serve as an annealing site for one or more linker-specific primers used in the invention methods as described herein.

For example, in a preferred embodiment according to the invention, the first oligonucleotide in the invention topoisomerase-adapted linker comprises a first oligonucleotide sequence

5'-TAGAAGGCACAGTCGAGGACTTATCCTAGCCTCTGAATACTTTC
AACAAAGTTA-3' (SEQ ID NO: 9) and is adapted at the 5' end with an affinity tag, such as biotin, or a detectable molecule, such as a fluorescent molecule, to aid in selection of a polynucleotide to which the topoisomerase has transferred the invention linker from a heterogeneous mixture of polynucleotides. The biotin affinity tag also prevents the 5' end of the oligo from being a substrate for topoisomerase-mediated end-joining. In addition, the biotin modification facilitates reverse-phase chromatography oligo purification, eliminating the need for more costly gel-purification. The biotin group may also be useful in purifying the linked product before PCR by binding the biotin-ligated product to streptavidin. Other purification methods are known to those skilled in the art.

The second oligonucleotide is preferably fully complementary to the first nucleotide in the linker duplex and is phosphorylated at the 5' end thereof to ensure that it will not be a substrate for Topo-mediated DNA end-joining. For use with the

first oligonucleotide having a sequence according to SEQ ID NO: 9, the second oligonucleotide in the linker can have the sequence:

5'-AGGGTGTAAGTTGTTGAAAGTATTCAGAGGCTAGGATAAGTCCT
CGACTGTGCCTTCT-3' (SEQ ID NO:10)

- 5 Optionally the second oligonucleotide further comprises a 3' polyadenylation tail to aid in removal of the second oligonucleotide from a heterogeneous mixture of polynucleotides, as in the oligonucleotide:

5'-AGGGTGTAAGTTGTTGAAAGTATTCAGAGGCTAGGATAAGTCCT
CGACTGTGCCTTCTAAAAAAAAAAAAAAAA-3' (SEQ ID NO: 3). The polyA
10 stretch has been shown to bind oligodT cellulose to aid in preparation of the linker.

For use with a linker containing the above described first and second oligonucleotides, in the invention methods, linker-specific primers having the sequence:

5'-TAGAAGGCACAGTCGAGGACTTATCCTA-3' (SEQ ID NO: 11),
15 which anneals to the 5' end of the second oligonucleotide in the linker, or

5'-GCCTCTGAATACTTTCAACAAGTTA-3' (SEQ ID NO: 12), which
anneals to the area 3' of the target site of SEQ ID NO:3 can be used. The latter of the
two linker-specific primers is particularly designed for use with the above described
linker in nested PCR as the internal linker-specific primer.

- 20 Those of skill in the art will appreciate that, in general, the oligonucleotide sequence of linker-specific primers used for linker-mediated PCR amplification according to the invention methods will be designed to hybridize to whatever particular second oligonucleotide is used in the duplex linker under the conditions used for conducting the linker-mediated PCR amplification. For example, stringent
25 conditions can be used, such as are commonly used for Taq polymerase mediated primer extension, as is known in the art and as further illustrated in the Examples herein.

In another embodiment according to the present invention, there are provided methods for isolating a target polynucleotide having a segment of unknown sequence when sequence of some portion of the target polynucleotide is known, e.g., an adjacent segment of the target polynucleotide. The invention linker-assisted isolation method comprises:

- (a) cutting the target polynucleotide with a restriction endonuclease that leaves a 3' overhang of at least 2 bases to obtain an overhang-digested polynucleotide segment,
- (b) dephosphorylating the 5' ends of the overhang-digested polynucleotide segment,
- (c) performing a Taq polymerase mediated primer extension of the 5'-dephosphorylated overhang-digested polynucleotide segment using a sequence-specific primer to create a duplex extension product having at the 3' ends a single A base overhang,
- (d) incubating the duplex extension product with an invention topoisomerase-adapted linker so as to selectively attach the duplex DNA of the linker to the extension product to form a linked extension product, and
- (e) isolating a second extension product produced by amplifying the linked extension product using the sequence-specific primer and a topoisomerase linker-specific primer.

The dephosphorylated, A-tailed polynucleotide segments produced in step (d) of the invention method are the only acceptor sites for ligation of the invention Topoisomerase I adapted linker, allowing for specific ligation of the linker to the 3' ends of the polynucleotide segments prior to linker-mediated PCR amplification thereof. Therefore, in the use of invention methods to determine the sequence of an unknown segment of a target oligonucleotide, it is important to use a restriction endonuclease in step (a) of the invention method that cuts the target polynucleotide in the segment of unknown sequence. The preferred phosphatase for dephosphorylating the 5' ends of the cut DNA is calf intestinal phosphatase.

Optionally, the invention method may further comprise the step of obtaining the sequence of the isolated second extension product. It is presently preferred that the second extension product be gel-purified using fast performance liquid chromatography (FPLC) prior to sequencing. Thus, the invention methods of linker-mediated isolation and sequencing can be used to sequence a nucleic acid having a large segment of unknown sequence without cloning of the second extension product into a vector. Alternatively, the PCR products obtained by the invention methods can be readily cloned into a T/A[®] cloning vector or TOPO-T/A[®] cloning vector (Invitrogen, San Diego, CA) for subsequent sequencing.

10 Preferably the nucleic acid sequence is cut to completion with the restriction endonuclease, and the Taq polymerase used in the Taq polymerase mediated primer extension of the overhang-digested nucleic acid sequence is a Taq I polymerase that leaves the 3' ends of the extension products with a single A base overhang at the 3' ends. After dephosphorylation of the 5' ends of the extension product, a duplex
15 extension product with 3' A base overhangs is incubated with the invention topoisomerase-adapted linker so that the enzyme transfers the linker (i.e. in a topoisomerase "religation reaction") to the duplex extension product at the 3' A base overhang ends thereof. Thus, a linked extension product having at the 3' ends thereof a segment of nucleic acid of known sequence (i.e. the invention linker) is provided.
20 In the subsequent round of linker-mediated PCR amplification, a linker-specific oligonucleotide primer is annealed to the linker segment of the linked extension product. A second round of PCR extension, preferably under stringent conditions to assure sequence fidelity in the extension product, provides multiple copies of an isolated oligonucleotide that contains the linker-specific primer at the 3' end thereof.

25 As those of skill in the art will understand from the above description of the preferred embodiment, if a different sequence-specific topoisomerase is employed that ligates to DNA having a different overhang or other target characteristic, then the polymerase used to prepare the segment of DNA to be amplified is selected to match the requirements of the topoisomerase, thereby causing ligation of the invention linker

only to the 3' end of the segment of DNA (i.e., the end of unknown sequence) to be amplified.

In the preferred embodiment of the invention linker-mediated amplification methods, the preferred topoisomerase-adapted linker is adapted by covalent
5 attachment of *Vaccinia* topoisomerase and the duplex DNA comprises the nucleic acid sequences of SEQ ID NO: 9 and SEQ ID NO: 10 as described in detail above. For invention methods employing this linker, the preferred linker-specific primers are those having the nucleic acid sequence of SEQ ID NO: 11 or SEQ ID NO: 12, or a combination thereof. The linker-specific primer of SEQ ID NO: 11 is preferred for
10 use in nested PCR amplification, since this primer is designed to anneal during PCR to a segment of the linker 3' of the segment to which the linker-specific primer of SEQ ID NO: 12 anneals. Hence it is referred to herein as the "internal linker-specific primer." This linker preferred for use in the practice of the present invention has a length of 58 bp to provide sufficient known sequence for two mutually exclusive
15 target primer annealing sites, but the linker used in accordance with the invention can be of any convenient length so long as it is at least about 15 bp in length (the minimum length of a primer used in PCR amplification being generally about 15 nucleotides). Those of skill in the art will appreciate that in other embodiments of the invention, wherein the oligonucleotides in the duplex linker have a different sequence
20 than in the exemplary preferred embodiment, a similar strategy will be employed wherein the oligonucleotides that make up the duplex linker are designed to provide a known segment of nucleic acid sequence that serves as one or more, preferably mutually exclusive, target sites for annealing of sequence-specific primers during linker-mediated PCR amplification procedures.

25 If there is one strong PCR product detected by electrophoresis on agarose gel, the product can be gel-purified, preferably using SNAP columns as is known in the art and described in the Examples herein. For sequencing, one of the internal sequence-specific primers or internal linker-specific primers can be used, if desired. If the PCR amplification produces multiple bands on the purification gel, or to confirm the
30 identity of the PCR product before sequencing, the initial amplification product can

be diluted, for example in a 1/1000 dilution, and used as template in a nested PCR as described herein using sequence-specific and linker-specific primers that are internal to those used in the initial PCR amplification. This nested PCR amplification preferably will produce a single PCR product that can be purified and sequenced.

5 The above described linker-mediated amplification methods are specifically designed for contiguous sequencing of genomic DNA of unknown sequence cloned in a vector, such as an artificial chromosome vector, wherein at least some sequence of the vector is known 5' to the clone contained in the vector, e.g., 5' to a cloning site therein. If a vector is used wherein the sequence of the vector is known in the area
10 5' to a cloning site, it is not necessary to know the sequence of any part of the cloned DNA or RNA nucleotide in order to obtain the sequence thereof. The invention method for linker-mediated amplification and sequencing of genomic DNA of unknown sequence contained in such a vector comprises:

- 15 (a) cutting the nucleotide sequence of the vector with a restriction endonuclease that leaves a 3' overhang of at least 2 bases to obtain an overhang-digested polynucleotide segment that contains an oligonucleotide of unknown sequence,
- (b) dephosphorylating the 5' ends of the overhang-digested polynucleotide segment,
- 20 (c) performing a Taq polymerase mediated primer extension of the 5'-dephosphorylated overhang-digested polynucleotide segment using a vector sequence-specific primer to create a duplex extension product having at the 3' ends a single A base overhang,
- (d) incubating the duplex extension product with an invention
25 topoisomerase-adapted linker so as to selectively join the duplex DNA contained in the linker to the extension product to form a linked extension product, and

- (e) isolating a second extension product produced by amplifying the linked extension product using the vector sequence-specific primer and a topoisomerase linker-specific primer.

The isolated second extension product can be directly sequenced to obtain the complete sequence thereof, preferably after FPLC gel purification, without first cloning the second extension product into a sequencing vector. Thus, in the case where the segment of genomic DNA of unknown sequence is contained within a vector, such as an artificial chromosome vector, the procedure used is similar to that described above wherein the sequence of a portion of a target polynucleotide is known, except that in the present embodiment the portion of target polynucleotide for which sequence is known is contained within the vector and the sequence-specific primer used (i.e., the 5' primer of the extension product) is designed to anneal to a segment of the vector of known sequence (and is therefore referred to herein as a "vector sequence-specific primer") rather than to a segment of the clone. Thus, in this embodiment of the invention, it is possible to isolate and sequence a target polynucleotide for which no sequence information is known.

For nested PCR in such a vector of known sequence, two or more such vector sequence-specific primers that anneal to "nested" segments of vector nucleic acid sequence are designed so that the vector specific primer and linker-specific primer used in the nested PCR amplification are both internal to that used in the primary round of linker-mediated PCR amplification.

The vector is generally an artificial chromosome vector and can be either a bacterial artificial chromosome vector (BAC) or a yeast artificial chromosome vector (YAC), such as a Lambda vector, as illustrated herein in Example 2 below. For example, a unique segment of about 1 kb from a BAC clone that contains a 170 kb human DNA insert has been successfully sequenced using the invention method.

In another embodiment according to the present invention, there are provided kits useful for performing the invention linker-mediated amplification methods. The invention kits comprise an invention topoisomerase-adapted linker and one or more

linker-specific primers that will anneal to the vector during PCR amplification. For example, the linker-specific primers can be fully complementary to a sequence of at least 15 contiguous nucleotides contained in the second oligonucleotide in the duplex linker. Preferably, the topoisomerase-adapted linker contained in the invention kit is the 58 bp duplex linker that comprises two oligonucleotide having the sequences of SEQ ID NO: 9 and SEQ ID NO: 10 and is adapted by covalent attachment of *Vaccinia* topoisomerase as described in detail above and the linker-specific primers contained in the kit are one or both of the oligonucleotides described herein as containing SEQ ID NO: 11 and/or SEQ ID NO: 12.

Optionally, the invention kit may further contain reagents useful for purification and direct sequencing of PCR products as described in the Examples herein. For example, reagents useful for performing NA-Iodide SNAP purification of polynucleotides may optionally be contained in the invention kits and/or reagents useful for performing Taq mediated PCR.

The invention kit is the first to use Topoisomerase I to attach a desired DNA to non-vector DNA, and it is the first to use a tagged Topoisomerase I to select a specific polynucleotide from a heterogeneous mixture of polynucleotides. The kit features a double-stranded DNA oligo that is covalently charged with *Vaccinia* Topoisomerase I (the topoisomerase-adapted linker) and include reagents for purification and direct sequencing of PCR products. The kit will likely be used for the analysis of genomic DNA elements such as exon-intron junctions and promoter regions. In addition, the kit should provide a rapid method to close gaps in sequence generated through shotgun sequencing.

The present invention methods are not limited to only unknown genomic DNA, and can be used to sequence any DNA or RNA under any situation. DNAs or RNAs of many different origins (e.g. viral, cDNA, mRNA) can be sequenced, not only for research or information gathering purposes, but also for other purposes, such as disease diagnosis and treatment, DNA testing, and forensic applications.

The invention will now be described in greater detail by reference to the following non-limiting examples.

The Examples herein describe a preferred method for manufacture of the invention pro-linker and topoisomerase-adapted linker. This method includes use of polynucleotide kinase (PNK) during the adaptation reaction to improve the efficiency of covalent attachment of Topoisomerase I to a desired DNA. Separation of Topoisomerase I/DNA complexes away from ATP, kinase, and free topoisomerase can then be accomplished using FPLC over a Sephacryl S-200 column, which is also shown to be an effective method to separate Topoisomerase I/DNA complexes away from ATP, kinase, and free topoisomerase.

EXAMPLE 1

PREPARATION OF A DOUBLE-STRANDED TOPOISOMERASE I ADAPTED LINKER

A. Oligos used to prepare pro-linker:

15 Linktop4bio

5'-TAGAAGGCACAGTCGAGGACTTATCCTAGCCTCTGAATACTTTC
AACAAAGTTACACCCTTATTCCGATAGTG (SEQ ID NO: 2). This oligo contains the *Vaccinia* Topoisomerase I cleavage site (in bold) and is biotinylated at the 5' end. The biotin prevents the 5' end of the oligo from being a substrate for Topo-mediated end-joining. In addition, the biotin modification facilitates reverse-phase chromatography oligo purification, eliminating the need for more costly gel purification. The biotin group may also be useful in purifying the linked product before PCR.

Linkbot4

25 5'-p-AGGGTGTAACCTTGTTGAAAGTATTCAGAGGCTAGGATAAGTCC
TCGACTGTGCCTTCTAAAAA (SEQ ID NO: 3). This oligo is phosphorylated at the 5' end to ensure that it will not be a substrate for

topoisomerase-mediated DNA end-joining. The polyA stretch has been shown to bind oligodT cellulose which may provide additional uses in separation of the linker.

Topo5

5'-CAACACTATCGGAATA (SEQ ID NO: 1). This oligo is modified by phosphorylation at the 5' end and anneals to Linktop4bio 3' of the phosphodiester bond cleaved by topoisomerase, to form the "leaving group". The oligos anneal as shown in Figure 1A to create the pro-linker; the structure of the invention linker after Topoisomerase I adaptation is shown in Figure 1B.

B. Annealing reaction:

- 10 1. In separate tubes, DNA oligos: biotinylated Linktop4bio, phosphorylated Linkbot4 and Topo5 (described in Section A above) were diluted to concentration 1×10^{-4} M in dH₂O.
2. The annealing reaction was set up as follows (final concentrations in brackets):
 15 22.2 µg Linktop4bio (1×10^{-5} M), 22.6 µg Linkbot4 (1×10^{-5} M), 5 µg Topo5 (2×10^{-5} M), 3.2 µl 5 M NaCl (160 mM), 1 µl 1M Tris-Cl, pH 8.0 (10 mM), and dH₂O to 100 µl total volume (total DNA concentration was 534 ng/µl). The mixture was heated to 94°C and slowly cooled to room temperature, then placed on ice.
3. A diluted stock mixture (5 ng/µl) of the annealed oligos to be used as a standard in
 20 the QC protocol below was prepared by combining 5 µl of the 534 ng/µl mixture with 529 µl tris-EDTA. Subsequently, the annealed oligos, which comprise the invention pro-linker were stored at minus 20°C.

C. Topoisomerase I adaptation of the pro-linker:

25 Topoisomerase I was adapted by mixing 35 µl of the annealed oligos, 87 µg *Vaccinia* Topoisomerase I, 40 units T4 polynucleotide kinase (NEB), 17.5 µl 10x polynucleotide kinase buffer (10x concentration = 700 mM Tris-HCl, pH 7.6; 100 mM MgCl₂; 50 mM dithio threitol, 4 µl 100 mM ATP, and dH₂O to a final volume of 175 µl. The mixture was incubated for 2 hours at 37°C followed by the addition of 3.5 µl 0.5 M ethylenediamine tetraacetic acid EDTA to chelate the

magnesium and inhibit possible nuclease activity during subsequent manipulations. In this reaction, the ratio of topoisomerase to topoisomerase cleavage sites in the DNA was approximately 5:1. 20 μ l of reaction mixture was saved out for subsequent analysis.

5 **D. Purification of Topoisomerase I adapted double-stranded oligonucleotides ("Topoisomerase-linker")**

FPLC gel-filtration purification was performed to separate the covalent Topoisomerase I/DNA complex from the other components of the Topo-adaptation mixture (PNK, ATP, ADP, free topoisomerase I, and free oligonucleotides).

10 Sephacryl S-200, which has a fractionation range for globular proteins of 5×10^3 to 2.5×10^5 Daltons and should exclude DNA >30 base pairs, was used as the column matrix. It was thought that this matrix would allow topoisomerase/DNA complexes to elute in or near the void volume, followed by free DNA, free protein, and then ATP. The column volume was ~24 mls and the column buffer was 200 mM NaCl,
15 1 mM EDTA, 10 mM Tris-Cl pH 8.0. The gel filtration column was run at 0.3 ml/minute throughout the procedure and eluate was continuously monitored by absorbance at 254 nm. The protocol for the FPLC run was as follows:

1. The column was equilibrated with 2.5 column volumes of column buffer.
2. The pump was stopped and 160 μ l adapted linker was slowly loaded to the column
20 with 1 ml syringe.
3. 200 μ l column buffer was used to chase the dead volume of the syringe and adapter into the column.
4. The pump was re-started and ~fifty 500 μ l fractions were collected.
5. Fractions were stored at 4°C until they were pooled.

25 **E. Pooling Fractions:**

1. The trace with optical density of 254 (OD₂₅₄) was used to determine which fractions comprise the initial elution peak (this peak contains the topoisomerase-linker).

2. The best of the fractions in the initial peak were pooled until the total volume was at least 1800 μ l. If fractions were eluted in \sim 500 μ l volume, at least the 4 gel-filtration fractions with the highest OD₂₅₄ reading were pooled.
3. Pooled fractions were stored at 4°C until dilution with storage buffer (see below).

5 Material eluted from the column in three distinct peaks, as indicated by the OD₂₅₄ results. Column fractions were analyzed by PAGE and silver staining, revealing a strong band in fractions representing a complex comprised of one topoisomerase molecule covalently attached to the topoisomerase-linker DNA (as shown in Figure 1B).

10 The FPLC fractions were also assayed for their ability to alter migration of a *Taq*-generated PCR product. A 300 bp PCR product was incubated with material from fraction 26 and fraction 46 (from the first and second elution peaks, respectively). Reaction mixtures contained the components indicated in Table 1 below in addition to 2 μ l 10x T/A PCR buffer (Invitrogen) and dH₂O to 20 μ l final
15 volume. After 20 min incubation at 37°C, proteinase K was added to the appropriate tubes and all reactions were incubated an additional 20 min at 37 °C. Reactions were loaded directly into a 4% E-gel. The results of these tests showed that migration of the PCR product was retarded after incubation with material from fraction 26, but retardation did not occur with fraction 46. The change in migration was due to a
20 covalent DNA addition to the 300 bp PCR product because the larger species was resistant to proteinase K. These results indicate that the topoisomerase-adapted DNA (topoisomerase-linker) found in fractions 24-31 can covalently attach to DNA that contains a 5'-hydroxyl end with a single A-base 3' overhang.

TABLE 1

| Reaction # | Fraction | 300 bp PCR product (10 ng/ μ l) | Proteinase K (5 mg/ml) |
|------------|------------------------------------|--|---------------------------|
| 1 | #26 – 13 μ l | 4 μ l | 1 μ l |
| 2 | #26 – 13 μ l | ---- | 1 μ l |
| 3 | #46 – 13 μ l | 4 μ l | 1 μ l |
| 4 | #46 – 13 μ l | ---- | 1 μ l |
| 5 | ----- | 4 μ l | ----- |
| 6 | #26 – 13 μ l | ----- | ----- |
| 7 | #46 – 13 μ l | ----- | ----- |
| 8 | 3 μ l Invitrogen 20 bp ladder | | |
| 9 | 3 μ l Invitrogen 100 bp ladder | | |

5 F. QC protocol:

QC consists of two 10% acrylamide gels and one functional assay (Linker-mediated amplification DNA produced by specific primer extension of Lambda DNA).

10 The acrylamide gels enable one to assess: 1) the amount of the topoisomerase-adapted oligonucleotide linker contained in the peak fractions (so that dilution factor can be determined), 2) the efficiency of the adaptation step, and 3) the ability of the linker to join to DNA ends created by Taq polymerase.

15 The functional assay was designed to demonstrate that the Linker in shipping buffer will enable Linker-mediated amplification DNA produced by specific primer extension of Lambda DNA.

1. QC-Gel One

Gel one enabled determination of a dilution factor for the purified topoisomerase-adapted oligonucleotide linker.

The pooled fractions contained a mixture of Topo-adapted and non-adapted DNA. In this experiment, the Topo-adapted DNA runs at ~80 bp while the non-adapted (uncut) DNA runs at ~100 bp. The Gel facilitates estimation of the concentration of cut DNA in the pooled fractions, thus determining the appropriate dilution factor for aliquotting. Reactions utilizing the various combinations of parameters shown in Table 2 below were prepared using a 10% TBE acrylamide gel (Novex) according to the manufacturer's instructions.

Briefly, the following procedure was followed:

1. Reagents were combined in the order of dH₂O, proteinase K (if used), then sample and incubated 30 minutes at 37° C. Then the 6x DNA loading dye was loaded onto the gel.
2. Samples were loaded onto the 10% TBE acrylamide gel and run until the blue dye (faster migrating of the two dyes reached the bottom of the gel.
3. The gel was opened and stained with ethidium bromide (30 min in mixture of 100 ml dH₂O + 100 µl 500 µg/ml ethidium bromide stain).
4. The gel was washed for 5 min in 100 ml dH₂O to remove the stain.

The gel was photographed using a UV light box.

TABLE 2

| reaction lane | sample (μ l) | 5 mg/ml ProteinaseK (μ l) | dH ₂ O | 6x DNA loading dye | DNA concentration in lane | dilution factor for linker |
|------------------|---|--------------------------------------|-------------------|--------------------------|---------------------------------|-------------------------------------|
| 1 | 1 μ l Invitrogen 20 bp ladder | 0 | 18 | 4 | | |
| 2 | 19 μ l pooled fractions | 0 | 0 | 4 | | |
| 3 | 19 μ l pooled fractions | 1 μ l | 0 | 4 | | |
| 4 | 1 μ l 5 μ g/ μ l annealed linker oligos | 0 | 19 | 4 | 0.208 | scrap prep |
| 5 | 2 μ l 5 μ g/ μ l annealed linker oligos | 0 | 17 | 4 | 0.416 | 5-fold |
| 6 | 3 μ l 5 μ g/ μ l annealed linker oligos | 0 | 16 | 4 | 0.625 | 7.5-fold |
| 7 | 19 μ l pooled fractions | 1 | 0 | 4 | | |
| 8 | 4 μ l 5 μ g/ μ l annealed linker oligos | 0 | 15 | 4 | 0.833 | 10-fold |
| 9 | 5 μ l 5 μ g/ μ l annealed linker oligos | 0 | 14 | 4 | 1.04 | 12.5- fold |

| reaction lane | sample (μ l) | 5 mg/ml ProteinaseK (μ l) | dH ₂ O | 6x DNA loading dye | DNA concentrati on in lane | dilution factor for linker |
|------------------|---|--------------------------------------|-------------------|--------------------------|----------------------------------|-------------------------------------|
| 10 | 6 μ l 5 μ g/ μ l annealed linker oligos | 0 | 13 | 4 | 1.25 | 15-fold |

2. Interpretation of Gel 1

If blue dye ran to the bottom of the gel, the 20 bp band of the ladder will have run off and the bottom band will be the 40 bp sized fraction). Lane 2 was loaded with a mixture of topoisomerase-adapted linker and non-adapted DNA. Topoisomerase is covalently attached to the adapted DNA, which prevents this DNA from running into the gel. The non-adapted DNA will run as ~100 base pair DNA. Lanes 3 and 7 were loaded with identically prepared samples and should look the same. Sample in these lanes was treated with Proteinase K to allow the topoisomerase-adapted DNA to run into the gel. Thus, lanes 3 and 7 should have an ~80 base pair band (from the topoisomerase-adapted DNA) that is not (or barely) seen in lane 2. Lanes 4, 5, 6, 8, 9 and 10 were "standard lanes" and were loaded with increasing amounts of ~100 bp DNA. Visual inspection was used to estimate which of the standard lanes contained roughly the same amount of DNA as is in the lower band (~80 bp) in lanes 3 and 7. The results in Table 2 above (columns 6 and 7) are then used to estimate the concentration of the adapted linker and to then determine the dilution factor.

3. QC-Gel Two

A second gel was run to show that the pooled peak fractions will join to a 400 bp PCR product and retard its migration. For this test reactions utilizing the various combinations of parameters shown in Table 3 below were prepared according to the following procedure:

1. Water, PCR buffer, PCR product diluted to 20 ng/μl in TE, and test sample was added to an appropriately labeled Eppendorf tube, mixed and allowed to incubate 30 min at 37°C.
2. Proteinase K was added to the tube, mixed, and allowed to incubate 30 min at 37°C.
3. Samples were loaded on 10% TBE acrylamide gel (Novex) and run until Blue dye (faster migrating dye) reached the bottom of the gel.
4. The gel was opened and stained with ethidium bromide (30 min in mixture of 100 mls dH₂O + 100 μl 500 μg/ml ethidium bromide stain).
5. The gel was washed for 5 min in 100 mls. dH₂O to remove the stain, and
6. The gel was photographed using a UV lightbox.

TABLE 3

| reaction/lane | test sample | 10x T/A PCR buffer | 20 ng/μl 400 bp LacZ PCR product | 5 mg/ml Proteinase K | dH ₂ O (μl) | 6x DNA loading Dye |
|---------------|-------------------------------|--------------------|----------------------------------|----------------------|------------------------|--------------------|
| 1 | 2 μl 100 bp Invitrogen ladder | 2 | 0 | 0 | 16 | 4 |
| 2 | 0 | 2 | 2 μl | 1 μl | 17 | 4 |
| 3 | 16 μl pooled fractions | 2 | 2 μl | 1 μl | 0 | 4 |
| 4 | 8 μl pooled fractions | 2 | 2 μl | 1 μl | 8 | 4 |
| 5 | 4 μl pooled fractions | 2 | 2 | 1 μl | 12 | 4 |
| 6 | 2 μl pooled fractions | 2 | 2 | 1 μl | 14 | 4 |

4. Interpretation of QC Gel Two

Gel Two should show that topoisomerase-linker causes PCR product to migrate more slowly in a concentration-dependent fashion. If QC Gel Two demonstrates that Linker does attach to PCR product and slows migration in a concentration-dependent fashion, proceed to Dilution of topoisomerase-linker as described below.

G. Dilution of Topoisomerase-linker:

An appropriate dilution factor was determined using Table 4 below, which shows reagent volumes for various possible dilutions. It is assumed that 1500 μ l of pooled fraction will be diluted (i.e., a 5-fold dilution will result in 7500 μ l linker in buffer, while a 10-fold dilution will result in 15,000 μ l linker in buffer). The buffer compositions used were as follows:

2x final wash: 60 mM Tris-Cl pH 7.4, 1 mM EDTA, 4 mM DTT, 0.2 mg/ml BSA and a few grains of phenol red.

10 Glycerol mix: 90% glycerol, 5 mM Tris-Cl pH 7.5, 0.1% Triton-X-100.

TABLE 4

| Dilution factor | μ l pooled fractions | 2x final wash (μ l) | Glycerol mix (μ l) | TE (μ l) | Total final volume (μ l) |
|-----------------|--------------------------|--------------------------|-------------------------|---------------|-------------------------------|
| 5 | 1500 | 1875 | 3750 | 375 | 7500 |
| 7.5 | 1500 | 2812.5 | 5625 | 1312.5 | 11,250 |
| 10 | 1500 | 3750 | 7500 | 2250 | 15,000 |
| 12.5 | 1500 | 4687.5 | 9375 | 3187.5 | 18,750 |
| 15 | 1500 | 5625 | 11250 | 4125 | 22,500 |

Using the above dilution table (Table 4), the final composition of the buffer for storing topoisomerase-linker is 45% Glycerol, 40 mM NaCl, 0.5 mM EDTA, 0.05% Triton-X-100, 20 mM Tris-Cl pH 7.49, 0.05% BSA, 1 mM DTT, and a few grains of phenol red.

EXAMPLE 2

This assay demonstrates that the diluted topoisomerase-linker joins to the A-overhang produced by a sequence-specific Taq-mediated primer extension of *Pst*I-digested Lambda DNA.

5 A. Sequence-specific primer extension of *Pst*I-cut Lambda DNA**1. Preparation of a polynucleotide segment from a vector clone**

A Lambda vector containing a DNA insert was cut to completion with *Pst*I and the cut DNA fragment was dephosphorylated using the following reaction mixture:

- 10 2 µg *Pst*I-cut Lambda DNA (Sigma)
 5 µl (5 units) calf intestinal phosphatase (CIP)(Roche)
 5 µl 10x dephosphorylation buffer (Roche)
 dH₂O to 50 µl

The reaction mixture was incubated 1h at 37°C to obtain a dephosphorylated fragment
15 containing insert DNA of unknown sequence

2. The *Pst*I cut dephosphorylated fragment was extracted with phenol and precipitated as follows:

1. Add 50 µl dH₂O
2. Add 50 µl phenol (Sigma) and vortex at full power for 10 seconds.
- 20 3. Centrifuge at full speed in microfuge (all centrifugation steps are performed at 4°C) for 1 minute.
4. Remove 90 µl supernatant to fresh Eppendorf flask and to this add 10 µl 5M Na-Acetate, 300 µl 100% ethanol, and 2 µl 20 mg/ml glycogen.
5. Vortex at full power for 10 seconds and place on dry ice for 15 minutes.
- 25 6. Centrifuge at full speed in microfuge for 15 minutes.
7. Pour off supernatant, add 500 µl 80% ethanol, mix by inverting several times, and centrifuge at full speed in microfuge for 15 minutes.

8. Pour off supernatant, carefully remove last traces of ethanol with 200 μ l pipettor (be sure to leave pellet undisturbed – it is better to leave 20 μ l ethanol behind than to disturb the pellet).
9. Dry pellet in speed-vacuum, then resuspend in 36 μ l dH₂O to obtain a product
- 5 having a DNA concentration will be ~50 ng/ μ l.

3. Primer extension was performed using PCR

- Two μ l (100 ng) resuspended DNA was combined with 2 μ l 10x T/A PCR buffer (Invitrogen), 1 μ l dNTP mixture (each base is at 2.5 mM in mixture), 50 ng primer 3512, 0.4 units *Taq* polymerase (Sigma), and sufficient dH₂O to obtain 20 μ l
- 10 of the mixture. Primer extension was then carried out using sequence-specific primer 3512 in a thermocycler according to the following profile: 4 min at 94°C, 1 min at 50°C, 20 min at 72°C, 10 min at 4°C. Primer 3512
- (5'- AACTCCGTGCAGCCGTACTG) (SEQ ID NO: 13) anneals at base 8565 in Lambda-DNA so that the extension would run until the *Pst*I site at bp 9615, where the
- 15 *Taq* polymerase produced a single A-base 3' overhang. The extension reaction mixture was stored at -20°C until used in the linking reaction (step 4 below).

4. Attachment of topoisomerase-linker to PCR product (extended Lambda DNA)

- The linking reaction mixture contained 1 μ l of topoisomerase-linker prepared
- 20 as described in Example 1 above, 1 μ l of extended Lambda DNA (see Section 3 above), 1 μ l 10x T/A PCR buffer (Invitrogen), and 7 μ l dH₂O. The reaction mixture was incubated 30 min at 37°C, then stored at -20°C until used as template for the linker-specific amplification reaction (see Section 5 directly below). The topoisomerase-linker in this reaction has been diluted to final shipping composition.

5. Topoisomerase Linker-specific PCR amplification

Linker-specific PCR amplification was performed utilizing the following primers:

Primer 2966 5'-GCACTGGAGAAGCATGACACCGGG-3' (SEQ ID
5 NO: 14), and
LinkAmp4 5'-TAGAAGGCACAGTCGAGGACTTATCCTA-3' (SEQ ID
NO: 11).

Primer 2966 anneals at base 8588 in lambda DNA, and primer LinkAmp4
10 anneals to the 5' region of the double stranded oligonucleotide topoisomerase linker
that has been adapted by covalent attachment of Topoisomerase I (prepared as
described above).

The PCR reaction was set up on ice as follows: 1 µl of linking reaction
mixture as prepared in Section 4 above, 50 ng primer 2966, 50 ng primer LinkAmp4,
15 1 µl 10 mM dNTP mixture (mixture contains 2.5 mM each dNTP), 2 µl 10x T/A PCR
buffer (Invitrogen), 2 units Taq polymerase (Sigma), and dH₂O to 20 µl. Then the
reaction mixture was cycled in a MJ research thermocycler according to the following
conditions: After a denaturing step of 4 min at 94°C, cycle 25 times (30" at 94°C, 30"
at 60°C, 1 min at 72°C), followed by 4 min at 72°C. The product of the PCR reaction
20 was stored at -20°C until used for agarose gel electrophoresis.

6. Analysis

For analysis, 10 µl of the PCR product was run on 2% electrophoresis-gel
alongside 4 µl Invitrogen mixed DNA ladder (100 bp and 1000 bp mixture). The
presence of 1 kb product indicated the topoisomerase-adapted linker was functional.

EXAMPLE 3**A. BAC DNA primer extension:**

The bacterial artificial chromosome used here is BAC 327L24. This artificial chromosome vector contains a 170 kb human DNA insert cloned into the *EcoRI* site of the pBAC3.6e backbone. DNA digestion cocktail: 1.7 µg DNA, 40 units *PstI* (NEB), 4 µl NEB restriction buffer 3, dH₂O to 40 µl final volume. This mixture was incubated at 37°C for 2 hr and then the enzyme was heat inactivated by 20 min incubation at 80°C. Subsequently, 4 µl calf intestine phosphatase (CIP) (Roche), 6 µl 10x phosphatase buffer (Roche), and 50 µl dH₂O were added and the mixture was incubated 1 hr at 37°C, then phenol extracted, precipitated, washed with 80% ethanol, dried, and resuspended to a concentration of 40 ng/µl (assuming complete DNA recovery).

120 ng of this cut and dephosphorylated DNA was used as template in a primer extension reaction performed as described above for Lambda DNA except that primer sequence-specific to the bacterial artificial chromosome vector were used instead of the Lambda-specific primer.

Primer bac3.6F3 (5'-TCTGTCCTTTTACAGCCAGTAGTG-3'; SEQ ID NO: 15) anneals to the pBac3.6e vector backbone and extends from bp 9597 (numbered as if the BAC-backbone does not contain an insert) to *PstI* site at bp 8536. Primer bac3.6F2 (5'- AGCGAGGAAGCACCAGGGAACA-3'; SEQ ID NO: 16) extends from bp 9536, and primer bac3.6F4 (5'-TTATATATTCTGCTTACACACGAT-3'; SEQ ID NO: 17) anneals just downstream of primer bac3.6F2.

B. Addition of Topoisomerase-linker to Taq-extended Lambda or BAC clone DNA:

Linking reaction mixture contained 1 µl of topoisomerase-linker (prepared as described in Example 1 above), 1 µl of the extended BAC DNA (see above), 1 µl 10x T/A PCR buffer (Invitrogen, San Diego, CA), and 7 µl dH₂O. Reactions were incubated 30 min at 37°C, then stored at -20°C until used as templates for topoisomerase-adapted linker-mediated amplification reactions (see directly below).

C. Topoisomerase linker-specific Primers used for Topoisomerase Linker-mediated amplification:

LinkAmp4 (5'-AGGCACAGTCGAGGACTTATCCTA-3'; SEQ ID NO: 18) anneals to the 5' region of the invention topoisomerase-linker prepared as described in Example 1 above, and is used for the initial amplification. It does not produce a product in single-primer PCR reactions using BAC DNA as a template. LinkAmp5 (5'-GCCTCTGAATACTTTCAACAAGTTA-3'; SEQ ID NO: 19) anneals to the 3' region of the topoisomerase-linker and is used for a second, nested amplification (moving (i.e. "walking" closer to the target section of the DNA) and as a sequencing primer (See Figures 1A and 1B).

D. Topoisomerase Activity Assay:

Vaccinia Topoisomerase I nicks and relaxes supercoiled DNA. To assay for topoisomerase activity of the invention topoisomerase-adapted linker, supercoiled DNA was incubated with the sample of interest and then the mixture was subjected to agarose gel electrophoresis to determine whether the DNA migrated as supercoiled or relaxed DNA. The assay components do not include DNA containing 5'-hydroxyl ends with single A-base 3' overhangs. Because the assay does not include a substrate for the topoisomerase-linker, topoisomerase that is covalently attached to Linker DNA remains attached to the Linker. Thus, the assay detects only free topoisomerase. This assay shows that the mixture loaded on the column contains free topoisomerase activity, but the pooled peak topoisomerase-linker fractions do not. This indicates that the column effectively separates free topoisomerase from the covalent DNA/topoisomerase complex. Thus, the assay detects free topoisomerase, not enzyme that is covalently attached to the duplex linker construct.

E. Purification and sequencing of PCR products:

PCR products were electrophoresed through 1% agarose gels run in 1x TAE buffer. Desired products were excised with a razor blade and removed from the agarose plugs by three different protocols. 1) *Spin-clean* purification: Excised plugs were frozen, placed in a SNAP mini-prep column, and centrifuged at full-speed in a microfuge for 2 min. 50 μ l dH₂O was then added to the column and a second 2 min

spin was performed. The total volume eluted was ~150 μ l and the agarose was entirely retained on the column frit. 2) *Spin-clean-precipitation* purification: 100 μ l of the Spin-clean eluate was precipitated by addition of 10 μ l 3 M Na-acetate and 300 μ l 100% ethanol, followed by 15 min incubation on dry ice and then 15 min centrifugation at 4°C. The pellet was then washed with 80% ethanol, dried, and resuspended in 40 μ l dH₂O. 3) *Na-Iodide-SNAP* purification: Performed according to protocol in XL-PCR kit (Invitrogen, San Diego, CA), in which each PCR product is ultimately eluted in 40 μ l dH₂O.

Sequencing reactions for DNAs purified by the *Spin-clean* or the *Spin-clean-precipitation* method contained 19 μ l DNA and 1 μ l (7 ng) LinkAmp5 oligo. Sequencing reactions for DNAs purified by the *Na-Iodide-SNAP* purification method contained 9 μ l DNA, 10 μ l dH₂O, and 1 μ l (7 ng) LinkAmp5 oligo as a sequencing primer.

EXAMPLE 4

A. Kinase improves efficiency of adaptation of the linker duplex by Topoisomerase:

Vaccinia Topoisomerase I binds to duplex DNA and cleaves the phosphodiester backbone of one strand after a 5'-CCCTT in the scissile strand. The phospho-tyrosyl bond between the DNA and enzyme can subsequently be attacked by the 5' hydroxyl of the original cleaved strand, thus reversing the reaction and releasing the topoisomerase. Because topoisomerase catalyzes both cleavage and religation, the structures in Figures 1A and 1B will reach an equilibrium. Phosphorylation of the 5' end of the leaving group should therefore drive the reaction towards the cleaved product, increasing the efficiency of preparation of the topoisomerase-adapted linker.

This hypothesis was tested by comparing the cleavage products obtained after annealed oligos were incubated with Topoisomerase I in the presence or absence of T4 polynucleotide kinase (PNK). A duplex linker oligonucleotide was prepared by annealing T7 top

(5'-GACTCGTAAATACGACTCACTATAGGTATCCGTGTCGCCCTTATTCCGA
TAGTGAC-3'; SEQ ID NO: 20) to a 5'-phosphorylated T7 bottom
(5'-AGGGCGACACGGATACCTATAGTGAGTGAGTCGTATTACGAG
TCTAG-3'; SEQ ID NO: 21), and Topo5 (SEQ ID NO: 1) at a molar ratio of 1:1:5,
5 respectively, in a mixture containing 70 mM Tris-HCl pH 7.6, 10 mM MgCl₂,
5 mM DTT, 200 mM NaCl. Topoisomerase adaptation reactions were set up with or
without 20 units PNK as follows: 5 µl (600 ng) annealed oligos, 70 mM Tris-HCl pH
7.6, 10 mM MgCl₂, 5 mM DTT, 5 mM ATP, 2 µg *Vaccinia* Topoisomerase, dH₂O to
final volume of 20 µl. Each reaction was incubated 30 min at 37°C, then split into
10 two 10 µl aliquots. One aliquot from each adaptation was then treated with 2 µl 5
mg/ml proteinase K for 30 min at 37°C. Samples were loaded onto a 15% TBE/urea
gel (Novex) as follows: Lanes: 1) 300 ng annealed oligos; 2) 1 µg T7-top oligo alone;
3) 1 µg T7 bottom oligo alone; 4) Topo-adaptation reaction with PNK and treated
with proteinase K before loading (100 ng DNA); 5) Topo-adaptation reaction with
15 PNK, without proteinase K treatment (100 ng DNA); 6) Topo-adaptation reaction
without PNK, with proteinase K (100 ng DNA); 7) Topo-adaptation reaction without
PNK and without proteinase K treatment (100 ng DNA).

The results of these tests show that addition of PNK increases accumulation of
the cleaved product in lines 4 and 6. Although the oligos used to illustrate this
20 principle are not those used for preparation of the preferred topoisomerase-adapted
linker, similar results were obtained using the oligo used in preparation of the
preferred topoisomerase-adapted linker. Therefore, PNK was used in all subsequent
adaptations of the topoisomerase linker by covalent attachment of Topoisomerase I.

However, Topoisomerase I is a DNA binding protein as well as a protein that
25 covalently attaches to DNA. The results of the above tests also illustrate that DNA
samples must be treated with proteinase K before electrophoresis to allow the DNA to
migrate correctly during agarose-gel electrophoresis.

Fractions were further characterized by assaying for DNA unwinding
(Topoisomerase) activity. Because the assay does not include a substrate for the

topoisomerase-linker, topoisomerase that is covalently attached to Linker DNA remains attached to the linker. Thus, the assay detects only free topoisomerase. The Sephacryl S-200 chromatography should also separate the covalent DNA/topoisomerase complex from PNK (Topoisomerase I and PNK have similar molecular masses - 33 kd vs. 30 kd, respectively). Kinase activity in the fractions could not be tested because the column load contained a great deal of ATP that would compete with the radioactive ATP used in standard kinase assays. However, a similar S-200 sephacryl column was loaded with kinase and DNA (and no ATP) and FPLC fractionation was performed. When fractions were assayed for kinase activity, it was shown that the fractions 24 -31 corresponding to about 12 through 14 ml collected, free of kinase activity (data not shown). This result suggests that the FPLC gel-purification will separate the covalent Topoisomerase I/DNA complex from PNK.

B. Topoisomerase-linker enables Linker-mediated Amplification (LMA) of a *Taq*-mediated primer extension product:

In order to develop topoisomerase-linkers as a tool to facilitate PCR amplification of DNA fragments for which the sequence of only one end is known, it was necessary to demonstrate that the topoisomerase-linker will attach to the product of a *Taq*-mediated primer extension, and that the attached DNA will serve as a primer-binding site for a subsequent PCR. When invention topoisomerase-linker was incubated with PCR extension products obtained using primers specific to known sequences of 50 kb Lambda phage DNA and 170 kb BAC DNA to obtain a template suitable for topoisomerase linker-mediated amplification, the templates were efficiently amplified using one Linker-specific primer and one primer fully complementary to a second, internal primer binding site on the extension product from the first round of PCR amplification. A contaminating band at ~400 bp in lanes 1 and 2 of the gel was not present when a different template-specific primer is used (lanes 3 and 4), indicating that results of an initial PCR can be influenced by primer design.

A second round of nested Taq-mediated PCR amplification was performed using as template the extension product from the first round of PCR amplification. For nested PCR a nested set of linker-specific and sequence-specific primers that are internal to those used in the first PCR extension were used to selectively amplify a shorter segment containing unknown sequence using the initial PCR products as a template. For nested PCR amplification, unpurified product from the reactions in lanes 1 and 2 was diluted 1/1000, amplified using nested primers, and the product was loaded in lanes 7 and 8, respectively. The nested PCR gave a robust product free of the 400 bp contaminating band. Subsequent sequence analysis confirmed the identity of the amplified DNA. Thus, the invention topoisomerase-linkers can be used to isolate DNA of unknown sequence and length and nested PCR can be used to confirm the identity of isolated amplification products so obtained.

C. PCR products generated by linker-mediated amplification can be sequenced directly:

The PCR products obtained as described above were purified from agarose by the *Spin-clean* method. Good quality sequence was obtained from 9 templates of about 500 bp in length purified by the *Spin-clean* method, while none of the *Spin-clean-precipitation* DNAs gave any sequence. It is likely that the poor results with the precipitation method were the result of DNA loss during the precipitation.

In order to compare *Spin-clean* sequencing results with *Na-Iodide-SNAP* results, the same 9 templates were reamplified and the products purified according to the *Na-Iodide-SNAP* method. These templates gave excellent sequence, with most reads >550 bases. The primer LinkAmp5 was used for all sequencing reactions, demonstrating that it can be used for PCR products generated using either LinkAmp4 or LinkAmp5. Sequencing results are summarized in Table 2 (the number in the "Bases Read" column indicates the base at which the sequence became difficult to call by visual inspection).

D. Conclusion:

The results described above also demonstrate that the topoisomerase-adapted linker covalently attached to an A-tailed PCR product, and the linker enabled PCR amplification of templates generated through gene-specific primer extension of
5 Lambda and BAC DNA. It is also demonstrated that a nested PCR performed with the internal linker-specific primer (LinkAmp5) selectively amplified the desired DNA from the mixture of DNA species contained within an initial Linker-mediated-amplification product. In addition, it is shown that PCR products generated through
10 Linker-mediated-amplification could be purified using SNAP columns and then sequenced directly using the LinkAmp5 primer (Table 1). DNA purified using the *Na-Iodide-SNAP* method gave the highest quality sequencing results.

While the invention has been described in detail with reference to certain preferred embodiments thereof, it will be understood that modifications and variations are within the spirit and scope of that which is described and claimed.

WHAT IS CLAIMED IS:

1. A double stranded topoisomerase-adapted linker comprising a duplex oligonucleotide linker, wherein a first oligonucleotide is annealed to a second oligonucleotide that is phosphorylated at the 5' end thereof, and wherein the linker
5 further comprises a compatible site-specific topoisomerase enzyme covalently attached to the first oligonucleotide at a single base 3' T overhang.
2. A topoisomerase-adapted linker according to claim 1, wherein the site-specific topoisomerase is a Topoisomerase I.
3. A topoisomerase-adapted linker according to claim 1, wherein the site-
10 specific topoisomerase is a *Vaccinia* topoisomerase and the cleavage site comprises SEQ ID NO: 4.
4. A topoisomerase-adapted linker according to claim 1, wherein the first oligonucleotide comprises an oligonucleotide sequence according to SEQ ID NO: 9 and the second oligonucleotide is complementary to the first nucleotide and is
15 phosphorylated at the 5' end thereof.
5. A topoisomerase linker according to claim 4, wherein the first oligonucleotide comprises an oligonucleotide sequence according to SEQ ID NO: 9 and the second oligonucleotide comprises an oligonucleotide sequence according to SEQ ID NO: 10.
- 20 6. A topoisomerase linker according to claim 1, wherein the first oligonucleotide is modified by attachment of biotin at the 5' end thereof.
7. A topoisomerase linker according to claim 1, wherein the second oligonucleotide further includes a poly-A tail at the 3' end thereof.

8. A topoisomerase linker-specific oligonucleotide having a sequence according to SEQ ID NO: 11.

9. A topoisomerase linker-specific oligonucleotide having a sequence according to SEQ ID NO: 12.

5 10. A method for isolating a target polynucleotide having a segment of unknown sequence when sequence of 3' nucleic acid is known, said method comprising:

10 (a) cutting the target polynucleotide with a restriction endonuclease that leaves a 3' overhang of at least 2 bases to obtain an overhang-digested polynucleotide segment,

 (b) dephosphorylating the 5' ends of the overhang-digested polynucleotide segment,

15 (c) performing a Taq polymerase mediated primer extension of the 5'-dephosphorylated overhang-digested polynucleotide segment using a sequence-specific primer to create a duplex extension product having at the 3' ends a single A base overhang,

20 (d) incubating the duplex extension product with a topoisomerase-adapted linker according to claim 1 so as to selectively attach the duplex DNA of the linker to the extension product to form a linked extension product, and

 (e) isolating a second extension product produced by amplifying the linked extension product using the sequence-specific primer and a topoisomerase linker-specific primer.

11. The method according to claim 10, wherein the method further comprises obtaining the sequence of the isolated second extension product.

12. The method according to claim 11, wherein the second extension product is FPLC gel-purified prior to sequencing.

5 13. The method according to claim 11, wherein the sequence is obtained without cloning of the second extension product into a vector.

14. The method according to claim 10, wherein the Taq polymerase is Taq I polymerase.

10 15. The method according to claim 10, wherein the sequence-specific topoisomerase is a *Vaccinia* topoisomerase enzyme.

16. The method according to claim 10, wherein the topoisomerase linker comprises a covalently attached *Vaccinia* topoisomerase I and binds only to DNA containing a single 3' A base overhang.

15 17. The method according to claim 10, wherein the topoisomerase linker is the topoisomerase linker according to claim 4 and the linker-specific primer is a nucleotide having the sequence of SEQ ID NO: 11.

18. The method according to claim 10, wherein the topoisomerase linker is the topoisomerase linker according to claim 4 and the linker-specific primer is a nucleotide having the sequence of SEQ ID NO: 12.

20 19. The method according to claim 10, wherein the target nucleotide sequence is DNA.

20. The method according to claim 10, wherein the target nucleotide sequence is RNA.

21. A method for contiguous sequencing of genomic DNA having a segment of unknown sequence in a vector, said method comprising:

- 5 (a) cutting the nucleotide sequence of the vector with a restriction endonuclease that leaves a 3' overhang of at least 2 bases to obtain an overhang-digested polynucleotide segment that contains an oligonucleotide of unknown sequence,
- (b) dephosphorylating the 5' ends of the overhang-digested
10 polynucleotide segment,
- (c) performing a Taq polymerase mediated primer extension of the 5'-dephosphorylated overhang-digested polynucleotide segment using a vector sequence-specific primer to create a duplex extension product having at the 3' ends a single A base overhang,
- 15 (d) incubating the duplex extension product with a topoisomerase-adapted linker according to claim 1 so as to selectively join the duplex DNA contained in the linker to the extension product to form a linked extension product, and
- (e) isolating a second extension product produced by amplifying the
20 linked extension product using the vector sequence-specific primer and a topoisomerase linker-specific primer.

22. The method according to claim 21 further comprising obtaining a nested extension product by nested Taq-mediated PCR amplification of the second extension product using a set of sequence-specific and topoisomerase linker-specific primers
25 that are internal to the primers used in step (d) above.

23. The method according to claim 21, wherein the overhang-digested DNA is from about 1 kilobase to about 50 kilobases in length.

24. The method according to claim 21, wherein the vector is an artificial chromosome vector.

5 25. The method according to claim 24, wherein the artificial chromosome vector is a bacterial artificial chromosome vector.

26. The method according to claim 24, wherein the artificial chromosome vector is a yeast artificial chromosome vector.

10 27. The method according to claim 26, wherein the yeast artificial chromosome vector is a Lambda vector.

28. The method according to claim 27, wherein the restriction endonuclease is PstI.

15 29. The method according to claim 21, wherein the topoisomerase linker is a topoisomerase linker according to claim 4, and wherein the topoisomerase linker-specific primer has a nucleic acid sequence according to SEQ ID NO: 11.

30. The method according to claim 21, wherein the method further comprises sequencing the isolated second extension product.

31. A kit for sequencing contiguous sequencing of genomic DNA that is cloned in a vector, said kit comprising:

- (a) topoisomerase-adapted linker according to claim 1, and
- (b) a topoisomerase linker-specific oligonucleotide,

5 wherein the linker-specific oligonucleotide hybridizes with the duplex DNA of the linker under stringent conditions.

32. The kit according to claim 31, further comprising:

- (c) a PstI-cut Lambda vector and
 - (d) a nucleotide primer that anneals to the PstI-cut Lambda vector,
- 10 wherein the genomic DNA is cloned in the Lambda vector.

33. A method for amplifying by the polymerase chain reaction (PCR) a portion of a target nucleic acid sequence, said method comprising:

- (a) cutting the DNA with a restriction endonuclease that leaves a 3' overhang of at least 2 bases to obtain an overhang-digested DNA,
 - (b) removing the 5' phosphate groups from the overhang-digested DNA,
 - (c) performing a Taq polymerase-mediated primer extension of the overhang-digested DNA using a sequence-specific primer to create a duplex extension product having at the 3' ends a single A base overhang,
- 15
- 20

(d) incubating the duplex extension product with a topoisomerase-adapted linker according to claim 1 so as to covalently attach the duplex DNA of the linker to the extension product to form a linked extension product, and

5 (e) performing PCR amplification of the linked extension product using the sequence-specific primer and a topoisomerase linker-specific primer.

34. The method according to claim 33, wherein said amplification is under conditions of sufficient stringency so that specific amplification occurs.

10 35. The method of claim 34, further comprising sequencing the amplification product of step (e), thereby determining the sequence of a portion of said target nucleic acid sequence.

36. The method of claim 33, wherein the target nucleic acid is DNA.

37. The method of claim 33, wherein the target nucleic acid is RNA.

15 38. The method of claim 35, wherein the target nucleic acid is DNA.

39. The method of claim 35, wherein the target nucleic acid is RNA.

FIGURE 1A

| Linktop4bio | LinkAmp4 | LinkAmp5 | Topo-site |
|-----------------------------------|---|---|---|
| biotin- 5' aaaaaaaaaaaaaaaa | tagaaggcacagtcgaggacttatacctagcctctgaatactttcaacaagttacaccccttattccgatatgtg aaaaaaaaaaaaaaaa | tagaaggcacagtcgaggacttatacctagcctctgaatactttcaacaagttacaccccttattccgatatgtg aaaaaaaaaaaaaaaa | 3' attccgatatgtg aaaaaaaaaaaaaaaa |
| Linkbot4 | | | ↑ break Topo5 |

FIGURE 1B

| | | | |
|-----------------------------------|--|--|-----------------------------------|
| biotin- 5' aaaaaaaaaaaaaaaa | tagaaggcacagtcgaggacttatacctagcctctgaatactttcaacaagttacacccctt aaaaaaaaaaaaaaaa | tagaaggcacagtcgaggacttatacctagcctctgaatactttcaacaagttacacccctt aaaaaaaaaaaaaaaa | attccgatatgtg aaaaaaaaaaaaaaaa |
| | | | Leaving group → |

SEQUENCE LISTING

<110> INVITROGEN CORPORATION
HEYMAN, John
SZALAI, Gabor

<120> TOPOISOMERASE LINKER-MEDIATED AMPLIFICATION METHODS

<130> INVIT1270WO

<140> Herewith

<141> 2001-02-23

<160> 21

<170> PatentIn version 3.0

<210> 1

<211> 16

<212> DNA

<213> ARTIFICIAL

<220>

<223> oligonucleotide Topo5

<400> 1

caacactatc ggaata

16

<210> 2

<211> 72

<212> DNA

<213> ARTIFICIAL

<220>

<223> oligonucleotide linktop4bio

<400> 2

tagaaggcac agtcgaggac ttatcctagc ctctgaatac ttcaacaag ttacaccctt

60

attccgatag tg

72

<210> 3

<211> 72

<212> DNA

<213> ARTIFICIAL

<220>

<223> oligonucleotide - linkbot4

<400> 3

agggtgtaac ttgttgaaag tattcagagg ctaggataag tctcgactg tgccttctaa

60

aaaaaaaaaa aa

72

<210> 4
<211> 5
<212> DNA
<213> ARTIFICIAL

<220>
<223> consensus pentapyrimidine element

<400> 4
ycctt

5

<210> 5
<211> 12
<212> DNA
<213> ARTIFICIAL

<220>
<223> example of alternative Vaccinia topoisomerase

<400> 5
gcccttattc cc

12

<210> 6
<211> 12
<212> DNA
<213> ARTIFICIAL

<220>
<223> example of alternative Vaccinia topoisomerase

<400> 6
tcgcccttat tc

12

<210> 7
<211> 12
<212> DNA
<213> ARTIFICIAL

<220>
<223> example of alternative Vaccinia topoisomerase

<400> 7
tgtcgccctt at

12

<210> 8
<211> 12
<212> DNA
<213> ARTIFICIAL

<220>
<223> example of alternative Vaccinia topoisomerase

<400> 8

gtgtcgcct ta

12

<210> 9
<211> 53
<212> DNA
<213> ARTIFICIAL

<220>
<223> first oligonucleotide in the linker

<400> 9
tagaaggcac agtcgaggac ttatcctagc ctctgaatac tttcaacaag tta

53

<210> 10
<211> 58
<212> DNA
<213> ARTIFICIAL

<220>
<223> second oligonucleotide in the linker

<400> 10
aggggtgaac ttgttgaaag tattcagagg ctaggataag tcctcgactg tgccttct

58

<210> 11
<211> 28
<212> DNA
<213> ARTIFICIAL

<220>
<223> oligonucleotide that anneals to the 5' end of the second oligonucleotide in the linker

<400> 11
tagaaggcac agtcgaggac ttatccta

28

<210> 12
<211> 25
<212> DNA
<213> ARTIFICIAL

<220>
<223> oligonucleotide that anneals to the area 3' of the target site

<400> 12
gcctctgaat actttcaaca agtta

25

<210> 13
<211> 20
<212> DNA
<213> ARTIFICIAL

<220>

<223> primer 3512 anneals at base 8565 in Lambda DNA

<400> 13

aactccgtgc agccgtactg

20

<210> 14

<211> 24

<212> DNA

<213> ARTIFICIAL

<220>

<223> linker-specific PCR amplification

<400> 14

gcactggaga agcatgacac cggg

24

<210> 15

<211> 24

<212> DNA

<213> ARTIFICIAL

<220>

<223> primer bac3.6F3

<400> 15

tctgtccttt tacagccagt agtg

24

<210> 16

<211> 22

<212> DNA

<213> ARTIFICIAL

<220>

<223> primer bac3.6F2

<400> 16

agcgaggaag caccagggaa ca

22

<210> 17

<211> 24

<212> DNA

<213> ARTIFICIAL

<220>

<223> primer bac3.6F4

<400> 17

ttatatattc tgcttacaca cgat

24

<210> 18

<211> 24

<212> DNA
<213> ARTIFICIAL

<220>
<223> primer LinkAmp4

<400> 18
aggcacagtc gaggacttat ccta

24

<210> 19
<211> 25
<212> DNA
<213> ARTIFICIAL

<220>
<223> primer LinkAmp5

<400> 19
gcctctgaat actttcaaca agtta

25

<210> 20
<211> 56
<212> DNA
<213> ARTIFICIAL

<220>
<223> oligonucleotide, part of topoisomerase adaptation reactions

<400> 20
gactcgtaaa tacgactcac tataggtatc cgtgtcgccc ttattccgat agtgac

56

<210> 21
<211> 47
<212> DNA
<213> ARTIFICIAL

<220>
<223> oligonucleotide, part of topoisomerase adaptation reactions

<400> 21
agggcgacac ggatacctat agtgagtgag tcgtattacg agtctag

47

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/05745

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 15/64, C12P 19/34, C12N 15/00

US CL : 435/91.41, 91.1, 91.4, 91.5, 172.3

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/91.41, 91.1, 91.4, 91.5, 172.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|--|-----------------------|
| X | US 5,766,891 A (SHUMAN) 6 June 1998 (6.06.1998), see whole document. | 1-3 and 6 |
| A,P | US 6,238,884 B1 (SHORT et al) 29 May 2001 (29.03.2001), see whole document. | 1-39 |

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

15 June 2001 (15.06.2001)

Date of mailing of the international search report

17 JUL 2001

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Alexander H. Spiegler

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/05745

Continuation of B. FIELDS SEARCHED Item3:

Databases: WEST, Medline, Biosis, CAPlus, Embase, Biotechds, Sequence Search

Search Terms: topoisomerase, linker, PCR, biotin, Vaccinia, Taq polymerase, overhang, sequencing